DOCUMENT RESUME

ED 447 599 EA 030 774

AUTHOR Ross, Steven M.
TITLE How To Evaluate Comprehensive School Reform Models. Getting Better by Design, Volume 8.
INSTITUTION New American Schools Development Corp.
SPONS AGENCY Annenberg Foundation, St. Davids, PA.; Education Commission of the States, Denver, CO.
PUB DATE 2000-00-00
NOTE 25p.
AVAILABLE FROM New American Schools, 1560 Wilson Boulevard, Suite 901, Arlington, VA 22209. Tel: 703-908-9500; e-mail: info@nasdc.org; Web site: http://www.naschools.org.
PUB TYPE Guides - Non-Classroom (055)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Change; *Educational Methods; Elementary Secondary Education; Formative Evaluation; *Outcomes of Education; *Program Evaluation; Summative Evaluation
IDENTIFIERS *Reform Efforts

ABSTRACT

This guide is designed to help school district leaders promote comprehensive school reform (CSR) and to assist schools in evaluating the implementation and outcomes of their CSR models. The guide discusses why evaluation is critical and how to use it as a tool for continuous improvement in CSR models. Summative and formative evaluations are discussed. Among the key outcomes that formative evaluations consider are implementation progress, school climate, classroom teaching and learning activities, and teacher buy-in. Each of these areas is examined separately, and suggestions are made for the data collection instruments, procedures, analyses, and interpretation of results appropriate for each outcomes. The process of conducting evaluations for continuous improvement should yield a diagnostic report that forms the basis for discussion with school staff, the principal, the school leadership team, and appropriate parents, community representatives, and district administrators. A case study demonstrates how the Memphis City School District, Tennessee, is measuring outcomes on a regular basis. A glossary explains terms that may be unfamiliar to users. (Contains 6 figures and 11 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 447 599

New American Schools

How to Evaluate Comprehensive School Reform Models

Steven M. Ross

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Kovar

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

How can schools and school districts measure the progress and results of their comprehensive school reform designs? What types of evaluation methods, instruments, and procedures are available? How should different types of data be interpreted? Steven M. Ross explains how schools can and should use evaluation as a tool for data-driven decisionmaking and continuous improvement and provides a case study illustrating how the Memphis City Schools are measuring outcomes on a regular basis.

EA 030774

BEST COPY AVAILABLE

New American Schools

New American Schools (NAS) is a dynamic coalition of teachers, administrators, parents, community and business leaders, policymakers, and experts from across the country committed to improving achievement for all students by dramatically changing America's classrooms, schools, and school systems.

Unlike many reforms that are add-on programs or isolated projects, NAS designs aim to improve the whole school, from curricula and instruction to funding and community involvement. Recognizing that one size does not fit all schools and communities, NAS offers a choice of different designs — blueprints — for helping all students achieve at high levels. (For information on each design, turn to the inside back cover.)

New American Schools has clear and consistent goals:

- Establish supportive and assistance-oriented school systems.
- Develop school and teacher capacity to teach all students to high academic standards.
- Spend resources wisely with an eye to student results.
- Build broad and deep community support for education improvement and excellence.
- Make America's public schools places where all students excel.

New American Schools is results-oriented.

In a short period of time, NAS has generated impressive results. In many schools that are using a NAS design:

- students are producing higher-quality work, achieving at higher levels, and showing improvement on standardized tests and other measures of performance;
- discipline problems are down and student attendance and engagement are up;
- teacher enthusiasm and community involvement are on the rise; and
- student achievement is improving more quickly than conventional wisdom suggests is possible.

New American Schools helps partner districts restructure.

To overcome traditional barriers to school excellence, NAS provides focused assistance to its district partners in five key areas:

- rethinking school finance, including investment funding and resource reallocation strategies;
- revamping professional development infrastructures to support whole-school transformation;
- setting high academic standards and linked assessments;
- giving schools authority to make decisions about curriculum, staff, and spending and then holding schools accountable for results; and
- engaging parents and the public in school-improvement efforts.

New American Schools believes in shared accountability.

The foundation of NAS is a strong partnership built on shared responsibility for results. Clearly defined roles link partners to one another and to results. All stakeholders in a NAS community — teachers, administrators, district leaders, parents, and NAS Design Teams — are expected to take responsibility and to be held accountable for helping to improve student achievement. NAS partners also commit to regular and rigorous assessment of their performance, resulting in the sound business practice of continuous improvement. The RAND Corporation is the independent evaluator of the New American Schools effort.

© 2000 by New American Schools

How to Evaluate Comprehensive School Reform Models

Steven M. Ross

HOW TO USE THIS GUIDE

Schools across the country are dramatically changing the way they operate by implementing comprehensive school reform (CSR) models. These reforms are intended to make schools more efficient and successful in preparing students for the 21st century workforce. The CSR movement calls for schools and districts to focus their reform efforts on all aspects of school functioning — instruction, curriculum, governance, professional development, parental and community involvement, and support services. Today, federal programs such as Title I schoolwide projects and the Comprehensive School Reform Demonstration program ("Obey-Porter") directly promote schools' adoptions of CSR models.

This guide is designed to help school district leaders promote CSR and to assist schools in evaluating the implementation and outcomes of their CSR models. CSR evaluations support "data-driven" decisionmaking by schools, administrators, parents, and communities by providing information on how well a model is working; what changes the model is making in curriculum, instruction, organization, and results; and how the model can be improved over time. Another benefit of evaluation is the gathering of information about the relative success of different models, thus contributing to both local and

Steven M. Ross

Steven M. Ross, Ph.D., is professor of Educational Psychology and Research at the University of Memphis. A noted lecturer and researcher on evaluation, especially as it pertains to school reform strategies, Dr. Ross is the author of four textbooks and more than 125 papers for professional journals. He recently co-edited the book *Bold Plans for School Reform*, an analysis of the New American Schools designs, with Sam Stringfield and Lana Smith. Dr. Ross currently is collaborating with New American Schools, the Appalachian Educational Laboratory, and SERVE to help schools conduct formative evaluations for comprehensive school reform designs.

national understanding of which models work most effectively in particular contexts. With that in mind, this guide can be used to help school district leaders:

- increase understanding of **formative** and **summative** evaluations — and how to use each appropriately;
- provide information and resources to schools regarding evaluation methods, instruments, and procedures;
- make available to schools instrumentation that can be employed effectively and practically to collect data on CSR implementation and outcomes; and
- assist schools in interpreting and using evaluation results to reinforce program accountability and support continuous improvement.

The evaluation approach and specific instruments presented in this guide should be regarded only as guidelines or suggestions, not as prescribed means of assessing the effectiveness of CSR models. Much of the guide's content is based on the work of school districts and their efforts to develop and apply practical and informative evaluation strategies. The glossary at the end of the guide will help you understand any unfamiliar terms. As you read, please remember that each school district has unique evaluation needs — and that for your own community, those needs will dictate which approach is best. We encourage you to take advantage of the guide's strategies and suggestions and to use the many other quality evaluation resources available to schools and districts.

Development of this guide was supported by New American Schools (NAS), a private, nonprofit corporation established in 1991 to develop, promote, and support implementation of CSR models. NAS is dedicated to the fair and objective evaluation of its own efforts and is committed to helping districts and schools evaluate their CSR models.

CHAPTER 1: EVALUATION AND COMPREHENSIVE SCHOOL REFORM

Comprehensive school reform is not a quick-fix solution for schools and districts. The adoption of a CSR model is the first step in a long-term process that impacts every part of a school's operations. As an essential element of any CSR effort, schools and districts must evaluate closely and routinely both the implementation of models and the effects of those models on key indicators such as teaching, learning, parent involvement, and school climate. This guide aims to help schools and districts become more successful at designing and administering such evaluations. In the following pages, we will discuss why evaluation is critical and how to use it as a tool for continuous improvement in CSR models. We also will look at different forms of evaluation and how they can be employed effectively. Finally, we will present samples from existing data collection instruments that school and district staff should find useful as evaluation tools or prototypes.

Why Evaluate CSR Models?

Over the years, we have seen many once-promising educational programs come and go. Dissatisfied with past results, educators continually search for more effective strategies, enthusiastically experiment with those strategies and then, disappointingly, forget to do one crucial thing — systematically evaluate their success. But without a well-designed evaluation, how can school staffs and other stakeholders begin to understand how well the programs are working? How can they identify strengths and weaknesses, so that needed improvements can be made for the following year? Unfortunately, at many schools, teachers and principals — who have the greatest stake in and, possibly, bias toward a school's success — make these decisions subjectively, based on how a program "feels" or appears to be working. But for programs to improve and grow roots over time, valid evaluation data are needed to guide planning and implementation.

In the case of CSR, evaluation data can be used to help schools and districts accomplish a number of important goals:

- determine strengths and weaknesses of a newly implemented program;
- identify problems early in the implementation process and address them through a continuous improvement model;

- document early successes as positive feedback to school staffs and as supportive evidence for the continuance of the program;
- enable school staffs to base improvement planning on objective data; and
- formalize schools' accountability for the success of their comprehensive improvement program.

Summative and Formative Evaluations

What does "evaluation" mean to your school and district? Perhaps it implies different things, depending on the circumstances. There are two basic forms of evaluation — **summative** and **formative** — that we use in education. Summative evaluations judge final performance and, on that basis, help to determine rewards, sanctions, and future direction. When assessing CSR programs, summative evaluations look back and ask, *"How did the program do?"* Results then are used to guide decisions about whether the program should be expanded, maintained, or discontinued.

Unfortunately, by the time a summative evaluation is performed, it may be too late to improve a potentially good program that is not living up to expectations. Formative evaluations are more proactive and focus on implementation of a comprehensive school model and its early impacts on teachers, students, administrators, and other participants. This type of evaluation is aimed at monitoring and improving programs, asking questions like, *"How is the program doing?"* and *"How can the school better use the program to achieve goals?"* Among the key outcomes that formative evaluations consider are implementation progress, school climate, classroom teaching and learning activities, and teacher buy-in. Later in this guide, we will examine each of these areas separately and make suggestions for data collection instruments, procedures, analyses, and interpretation of results appropriate for each outcome.

An analogy that may help to sharpen the distinction between formative and summative evaluation can be made in thinking about cooking homemade soup for dinner guests. As you prepare the soup, you taste it and then adjust the ingredients ("more salt and pepper") based upon your impressions. This type of assessment is formative evaluation. That evening, you serve the finished soup, and, to your delight, the guests barrage you with requests for second helpings and the recipe. This type of assessment is summative evaluation, and the clear message here is to keep making that soup the same way!

Evaluation for Accountability or Continuous Improvement

Federal agencies, states, and school districts all have accountability requirements for the programs they fund. Frequently, these requirements impose benchmarks or standards specifying the outcomes programs must meet to be considered successful (e.g., student achievement levels, number of hours of teacher professional development, number of families served). The emphasis of these kinds of accountability requirements is on summative evaluation, even if assessments are made at various phases of program implementation.

Although accountability evaluations are important and pervasive in education, they are not the emphasis of this book. Rather, our focus is on using evaluation for continuous improvement (i.e., on formative evaluation). By systematically collecting data on how well their comprehensive reform programs are working, school staffs can make informed, reflective decisions on what adjustments to make. Remember the soup analogy? In education, particularly at the school level, the first time the soup is tasted frequently is at the meal! By then, it's too late to make changes. If the soup doesn't taste good, a good cook and potentially good recipe may be judged poorly.

The process of conducting evaluations for continuous improvement should yield a diagnostic report that forms the basis for discussion with the school staff, principal, school leadership team, and, where appropriate, parents, community representatives, and district administrators. Possible uses of this critical diagnostic tool include:

- grounding discussions of school improvement plans;
- measuring progress toward school goals;
- identifying professional development needs;
- clarifying instructional objectives provided by the district or state;
- identifying leadership, administrative, and other support needs; and
- strategizing with CSR teams on improving services.

CHAPTER 2: THE FORMATIVE EVALUATION PROCESS

To conduct a successful evaluation, it helps to have a basic model of evaluation procedures. One that we offer for illustrative purposes is displayed in Figure 1 on p. 13 (adapted from Ross & Morrison, 1995) and outlined below.

Step One: Designing an Evaluation Plan

The first step in developing a plan for a formative evaluation is to identify the evaluation's overall purpose and objectives. Let's say our purpose is to use the evaluation for continuous improvement of a CSR program. The evaluation objectives then would guide the entire process of determining data collection instruments, procedures, analyses, and the interpretation of results. The evaluation objectives often are framed as leading questions. Following are some evaluation questions that a school or district might find especially important to answer for continuously improving a CSR program.

Using multiple data sources enhances the quality and value of the evaluation. The combination of all these data gives schools and districts an accurate, in-depth look at the impact of CSR models on schools.

1. How effectively is the program being implemented? What are the strengths and weaknesses of the implementation process?
2. What are the effects of the program on classroom teaching and learning activities?
3. What are the effects of the program on student attendance, graduation rates, and achievement?
4. How supportive of the program are the key consumers (e.g., principal, teachers, parents, and students)?
5. How sufficient are the resources available to support the program? What additional resources are needed to improve program quality?
6. What are the effects of the program on the school environment (e.g., order, collaboration, expectations)?

Step Two: Designing the Methodology

Based on the evaluation's purpose and objectives, the next step is to design and implement the evaluation's methodology. First, the evaluator must identify the individuals and/or groups that will be surveyed, interviewed, or observed as part of the evaluation. In most CSR evaluations, interviewees will include teachers, district personnel, the principal, parents, and students. The evaluator also must determine how many interviewees from each group are needed (all or just a sample); if only a sample will be interviewed, the evaluator also must decide how the participants will be selected (e.g., randomly, by convenience, by targeting).

Next, the evaluator must determine what kind of data is needed to answer the guiding questions. If we were trying to answer the six sample questions from the previous section, we likely would need a variety of data. For example, question 4 focuses on support for the CSR program. This question could be answered with a combination of one-to-one interviews, focus groups (group interviews), and questionnaires administered to members of the selected participant groups. Question 6 looks at a CSR model's impact on school environment and might best be answered using a school climate survey. Getting answers to question 3, which focuses on student results, requires obtaining and measuring school data on the specified student outcomes.

Using multiple data sources enhances the quality and value of the evaluation. The combination of all these data gives schools and districts an accurate, in-depth look at the impact of CSR models on schools.

Step Three: Collecting the Data

After deciding whom to interview and what kind of data measurement instruments should be used, the evaluator must develop a plan for actually collecting the data. For example, how will participants be contacted and interviewed? A data collection plan usually includes the following tasks:

1. outlining the manner in which the data are to be collected;
2. designing and assigning the management of each part of the data collection process;
3. creating time lines for the accomplishment of various parts of the evaluation;
4. following the time lines and carrying out the data collection; and
5. storing the data in a safeguarded central location.

Step Four: Analyzing the Data and Interpreting Results

The data analysis phase begins with compiling, summarizing, and coding the data collected. Then, appropriate analyses should be identified and performed. Analyses are likely to include combinations of qualitative and quantitative methods. Based on the results, interpretations and conclusions can be made regarding each evaluation question.

Step Five: Reporting Results and Gathering Feedback

The final phase of the formal evaluation process is reporting the results and conclusions so that they can be used as information for decisionmaking and improvement of the CSR model. This usually involves writing a final report for specified audiences (e.g., the principal, teachers, parents, the superintendent, the school board, the Design Team); who the audiences are will depend on the purpose of the evaluation (e.g., whether it is intended to be formative or summative).

Step Six: Using Data for Continuous Improvement

Based on the evaluation data, the school leadership team (or entire staff, where appropriate) should use the findings as a basis for identifying needed improvements in both the implementation of the CSR model and in the evaluation itself. This ensures that the improvement process is continuous and dynamic.

CHAPTER 3: EVALUATION STRATEGIES AND TOOLS

It is one thing for principals and teachers to appreciate the need for program evaluation. It is quite another thing for principals and teachers to have the time and resources to perform such evaluations on their own. Recognizing this, the Center for Research in Educational Policy (CREP) at the University of Memphis has collaborated with the Appalachian Educational Laboratory (AEL) to develop a practical and affordable formative evaluation package to help schools assess their CSR programs. The package can be used in full or in part, depending on a school's individual interests and needs. What makes the package especially attractive to schools is that:

- instruments for surveys, interviews, school climate assessment, and classroom observations already are designed and available for use;
- survey and observation instruments are scannable, allowing schools to have results tabulated quickly and inexpensively; and
- data are analyzed and interpreted with recommendations in a final report written by the AEL/CREP staff, thus providing an external, unbiased review on the CSR program under evaluation.

In the following sections, we will examine various AEL/CREP instruments.

Benchmarking

How can school staff assess their progress in implementing a CSR design if they don't have a clear picture of what their design should "look like" with regard to classroom practice, student performance and products, professional development, and many other elements? Benchmarking helps schools and districts measure progress by providing key indicators of success for the most important elements of implementation.

The availability of pre-existing design benchmarks can be quite helpful in providing a foundation and starting point for the benchmarking process. For example, each of the New American Schools designs has a complete set of benchmarks to help schools measure success. Figure 2 on p. 14 shows a sample of the benchmarks used by Expeditionary Learning Outward Bound.

While it is essential for schools using a particular design to use the accompanying benchmarks, schools may need to develop additional benchmarks for a variety of

BEST COPY AVAILABLE

reasons. In some cases, standard components of a design may be implemented differently at an individual school (with input from the Design Team) in light of site conditions or district policies. In other cases, the standard benchmark may be too generic for teachers to use as a meaningful implementation guide. Further, schools may be accountable for supplementary programs developed in-house or mandated by the district or state.

When developing benchmarks, the school staff (usually a team of teacher leaders) should engage in meaningful discussion about what they are trying to accomplish, the desired progression in each area from the starting point to completion, and how the whole design should work to accomplish school goals. At the end of each year, the principal and faculty members should meet to identify progress on each indicator and what remains to be accomplished next year. In other words, the benchmarks should become a tool in the planning and program improvement process.

Regardless of the CSR design selection, school leadership teams that develop additional benchmarks also should:

1. identify all major components of their CSR program;
2. review and modify existing benchmarks using familiar descriptions and terminology and adjusting the benchmarks to fit site goals; and
3. work with the district and Design Teams to devise new benchmarks for areas not covered in the existing benchmarks.

Methodologies

School Climate Inventory (SCI)

Another valuable instrument is the School Climate Inventory (SCI), developed by CREP in 1989, which has been used for school-based improvement planning in schools nationwide. The inventory's seven scales (Order, Leadership, Environment, Involvement, Instruction, Expectations, and Collaboration) represent general factors associated with effective school management. Each scale contains seven statements, for a total of 49 statements — higher scores represent more positive results.

Figure 3 on p. 15 shows a sample output for Cloverdale Elementary School, a fictional test case. On each of the seven scales, Cloverdale experienced a small dip in 1998 and a noticeable gain in 1999 (the eighth set of data is a composite of the seven scales). The school's highest mean score is in Leadership, while its lowest is in Order. With these scores in hand, the Cloverdale staff can review these results to identify improvement goals and strategies for each coming year.

Comprehensive School Reform Teacher Questionnaire (CSRTQ)

The Comprehensive School Reform Teacher Questionnaire (CSRTQ) uses 28 factors rated on a five-point scale and two open-ended questions to assess teacher reactions to and experiences with a reform program. The questionnaire includes questions in professional development, resources, pedagogical change, and outcomes as a result of the design. Another section includes "tailored" items to assess progress toward school- and program-specific benchmark goals. A sample of teachers' perceptions from the CSRTQ ratings at Cloverdale Elementary School in spring 2000 is shown in Figure 4 on p. 16. These responses indicate teachers' levels of understanding of and support for the school's CSR program. For example, Item 1 shows that 40 percent of teachers agreed that they had a thorough understanding of the school's CSR program. Item 3 shows that 80 percent of the teachers agreed that the elements of the CSR program have been integrated effectively to help them meet school improvement goals.

School Observation Measure (SOM)

The ultimate goal of CSR is improving student learning. But we must be careful not to put the cart before the horse and remember that the usual cause of improved learning is improved teaching.

The School Observation Measure (SOM) highlighted in Figure 5 on p. 17 was designed to provide a "snapshot" of the teaching and learning activities taking place throughout a school. SOM is not CSR design-specific; however, it focuses on instructional strategies that commonly are associated with the education reform movement and can answer questions such as: "How frequently is cooperative learning used?" "Is technology use increasing over time?" "How frequently are parents seen at the school?" "Is teacher coaching highly prevalent as a strategy?" The SOM criteria are comprised of 24 targeted events grouped into six categories:

- Instructional Orientation (e.g., direct instruction, team teaching);
- Classroom Organization (e.g., ability groups, multiage grouping);
- Instructional Strategies (e.g., teacher coaching, project-based learning);

- Student Activities (e.g., experiential learning, sustained reading);
- Technology Use (e.g., using computers for instructional delivery); and
- Assessment (e.g., performance assessment, student self-assessment).

SOM observers make between 10 and 12 short visits to randomly selected classrooms and take notes relative to each of the 24 events. At the conclusion of each visit, observers rate the frequency of each event using a five category rubric: (0) Not Observed, (1) Rarely, (2) Occasionally, (3) Frequently, and (4) Extensively. SOM also includes two summary items that address student engagement and the use of academically focused time. Figure 5 shows the scannable form that observers complete at the end of their final visit. Note the scoring rubric at the bottom of the form.

It is suggested that SOM visits be conducted between six and 10 times a year, thus providing cumulative impressions of 60–120 classes in that school. For example, a school with a design that emphasizes cooperative learning might be pleased to learn that across 10 visits, cooperative learning was observed "Extensively" during five of the visits and "Frequently" during the other five visits. The school also might be pleased to find that its previously infrequent use of a desired strategy (e.g., experiential learning) is increasing steadily from year to year. On the other hand, the school might be concerned that project-based learning, a strategy emphasized by its NAS design, was "Not Observed" during 60 percent of the visits. Using this information, school leaders then might decide to increase professional development support in that area.

Interviews and Focus Groups

Paper-and-pencil surveys are an excellent means of gathering surface impressions on a wide range of topics from a large sample of respondents. Their limitation is that they do not get in-depth perspectives on the reasons for particular feelings and suggestions for improvements. Therefore, formative evaluations must include interviews with key school stakeholders — the principal, teachers, parents, and students. The decision to use group or individual interviews depends on conditions and personal choice. Individual interviews may result in more honest responses (respondents may feel uncomfortable expressing their opinions in a group) but are far more time consuming. On the other hand, group interviews offer opportunities for respondents to discuss issues and opinions with each other, which can lead to richer, more detailed impressions.

Generally, a one-hour interview with the principal is conducted to determine his or her perceptions and role in design implementation at the school. A group interview of similar length is conducted with seven to nine randomly selected teachers. Questions should focus on their support for the design; adequacy of the implementation overall; adequacy of professional development, resources, and planning time; outcomes in terms of student learning, changes in classroom teaching, and parent involvement; and suggestions for improvement.

After all the data are collected, they should be analyzed and synthesized in a diagnostic report prepared specifically and confidentially for the school.

Questions for principals and teachers should be simple and direct:

- Describe your role in the implementation of the CSR design in your school.
- How is the implementation of the CSR program going this year?
- Have you encountered any new challenges this year? (for those who are in at least the second year of a CSR design)
- What elements of the CSR program do you feel are the most effective? Least effective?
- What additional resources have been needed to support the CSR program in your school?
- How would you describe teacher support for the CSR program in your school?
- If I were to visit classrooms in your schools, what would I see that would demonstrate the presence of a CSR design?

Depending on resources, time, and participant availability, interviews with parents and students also may be conducted. Parents might be asked about their familiarity with the CSR design and how they feel the design has impacted them and their children. Students might be asked to describe how the design affects them, what they like and don't like about it, and suggestions for improving the design and the school in general.

BEST COPY AVAILABLE

Diagnostic Report

After all the data are collected, they should be analyzed and synthesized in a diagnostic report prepared specifically and confidentially for the school. This report should provide summary data from each of the instruments and document the current status of each of the benchmarks and school climate inventory scales. The report also should provide data on the degree to which different teaching strategies and classroom events were observed over the year; reactions from and suggestions by the principal, teachers, parents, and students; and achievement results from all available recent district or state tests. By discussing each set of results and concluding with recommendations for improving the CSR design during the following year, the report thus provides a data-driven foundation for school planning and serves as documentation of the school's progress in implementing the design and obtaining positive results. The report also can be used as the foundation for an annual report on progress prepared by the school for external stakeholders such as parents, community groups, and local businesses.

Evaluating CSR Program Effects: A Hierarchy of Strategies

The following evaluation strategies provide alternative approaches to measuring results, beginning with those strategies (Level I) likely to produce the most valid evidence.

Level I: Evaluating Program and Control Schools Using Standardized and Performance Measures. In this strategy, schools using CSR models first are matched with comparable control schools. Using stratified sampling procedures, students in the CSR and control schools are administered performance measures geared to district (or state) standards. Longitudinal analyses of both standardized and performance scores then are made, with appropriate disaggregation of data on variables such as gender, ethnicity, poverty, language proficiency, and number of years in school. If possible, value-added scores reflecting students' year-to-year gains are analyzed to control for these variables.

Level II: Evaluating Program and Control Schools Using Standardized Measures Only. This strategy is similar to that of Level I but omits the performance measures — most likely due to a lack of appropriate tests, or due to budget, personnel, testing logistics, or time restrictions.

Level III: Evaluating Program Schools Only Using Standardized and Performance Measures. This strategy is similar to that of Level I but does not include matched control schools. The absence of a control group obviously limits the degree to which CSR program effects on achievement can be inferred. However, suggestive evidence can be ascertained by examining over time students' (a) average gain relative to national, state, and/or district norms and (b) mastery of performance standards.

Level IV: Evaluating Program Schools Only Using Standardized Measures. This strategy is similar to that of Level III but does not use performance measures to indicate progress in achieving standards. Still, comparing standardized test scores to suitable national and/or local norms can suggest a CSR school's progress in raising achievement over time.

Evaluating CSR Program Effects: Parting Advice

Ultimately, for a CSR program to be successful, it must raise student achievement. But how fast and how noticeably can test scores be improved? The literature on school reform suggests that it may take up to six years to implement a CSR design thoroughly and successfully (Herman & Stringfield, 1995). However, key stakeholders in the reform effort, including the media and general public, may not be patient enough to wait that long for results. Therefore, to ease public concerns and strengthen and protect their CSR programs, district and school leaders should consider the following:

- **Tell your own story, or have someone else who understands less about the program or desired results tell it for you.** This means conducting formative evaluations and communicating the results to stakeholders. Recognize that, to preserve validity and credibility in such evaluations, you should use external evaluators to collect and analyze the data and produce the report. Still, those evaluators should work with you to tell the story in an appropriate and fair manner.
- **When results are positive (e.g., in student achievement or in the use of teaching strategies), use such information to strengthen both internal and external support for your CSR program.** In other words, get the word out both through the external evaluator and through your own publications. At the same time, be honest and open about needed improvements.
- **When results are negative, use them as a basis for continuous program improvement.** Remember to use them as a baseline to show progress the next year. Negative results (e.g., teacher dissatisfaction or lack of immediate achievement gains) are common in the beginning stages of CSR, as schools face the challenges of implementing new models and reforms. You should

be knowledgeable about those negative results so you can interpret and communicate them appropriately.

- **Ensure that your school's CSR design addresses the curriculum and content standards assessed on district and state tests.**
- **Use multiple measures of achievement to assess program effects.** If the program is effective, learning will be improved, but gains may show up more readily on certain types of tests than on others. If you don't provide alternative measures, someone else may pick the most available measure (usually the state-mandated multiple-choice test) and base your school's "story" solely on those results.
- **Educate stakeholders about the school reform process.** They need to understand that before achievement gains occur, teaching needs to be improved, which in turn depends on fully implementing the CSR design. Stakeholders should be educated to look for progress over time, not for immediate success.
- **Educate stakeholders about the differences between standardized multiple-choice tests and performance tests that are aligned with standards.** Also, educate them about the differences between improvement (value-added) scores and normative (percentile) scores. This will help them understand that different measures provide different information about student learning.
- **Educate stakeholders about the influences of student demographic variables, such as poverty, on test scores.** Comparing the median percentiles of diverse schools without considering such variables can be misleading. Looking at matched schools' performances (see the Level I and II designs on the previous page) and/or the performance of value-added schools creates a level playing field on which fair evaluation can take place.
- **Share your results with other schools that are using the same designs and with external Design Teams and developers.** Your experiences can be used to support continuous improvement of the CSR designs themselves.
- **Remember that evaluation needs to be ongoing.** School restructuring, like a flower garden, can prosper year after year with the proper monitoring and care. Without such care and evaluation in CSR schools, ineffective, more traditional structures and methods can creep back into school operations.

A CSR CASE STUDY: CONTINUOUS IMPROVEMENT IN MEMPHIS

In 1995, Memphis City Schools, led by Superintendent Gerry House, began a comprehensive school reform effort. Working in partnership with New American Schools, Memphis set a goal of implementing comprehensive school reform (CSR) designs in at least 30 percent of the district's schools within a five-year period. To prepare for implementation, school leadership teams reviewed literature and attended presentations on the NAS models. They brought this information back to their colleagues to determine faculty interest and then voted to select the design model that best fit each school's needs.

Thirty-four Memphis schools selected and implemented CSR models. From the beginning, systematic evaluation was planned as a key element of the district's overall reform effort. While it is still early to make summative evaluation judgments of the initiative's success, school leaders have been conducting an ongoing formative evaluation. The Center for Research in Educational Policy (CREP) at the University of Memphis has served as the district's partner in this evaluation effort.

Indicators of Program Processes and Outcomes

In planning the formative evaluation, CREP staff compiled the data sources they felt would not only address the district's specific evaluation questions, but that also would be useful for decisionmaking and practical to obtain.

The following are examples of the types of leading questions that helped CREP guide its work in Memphis.

- What activities and events need to be put in place to implement the selected design?
- How supportive are teachers of the design implementation? How do they view their preparation, activities, and roles? What is their impression of results?
- How supportive is the principal of the design? How does he/she perceive teacher buy-in, parent involvement, results, and the strengths and weaknesses of the design implementation?
- How is the design affecting the school in terms of its leadership, environment, organization, and collaborative spirit? Is this school a positive and safe place for teaching, learning, and visiting?
- How is the design impacting teaching and learning activities in the classroom?

BEST COPY AVAILABLE

- How is the design affecting student achievement and attendance and graduation rates?

Using existing and newly developed instruments, CREP collected data related to each of these questions during the school year. At the end of the year, results were reported individually for each school and collectively for the district. All reports gave the evaluators' impressions of the strengths and weaknesses of the design implementation and recommendations for improvement. As a result, the schools and the district had valuable, objective information for the next year's planning. Major results are summarized on p. 11.

The Memphis Evaluation Strategy: A Value-Added Analysis

Whether an evaluation is formative or summative, student achievement is the outcome of greatest interest in judging program effectiveness. The question all stakeholders want answered is: Are students learning more and performing better as a result of the CSR design?

To address this question in Memphis, the evaluation included studies on how student achievement at elementary schools implementing CSR models (known as "restructured schools") compared to achievement at matched control schools and at all other Memphis elementary schools (Ross, Sanders, Wright, & Stringfield, 1998; Ross, Wang, Sanders, Wright, & Stringfield, 1999). Data for such comparisons were derived from scores on the TerraNova (the state-mandated achievement test based on the Comprehensive Tests of Basic Skills — 5th edition) in five subjects (math, reading, language, science, and social studies) over a four-year period, shown in Figure 6 on p. 18.

These analyses represent a significant breakthrough in practices for assessing student achievement through examination of improvement in students' year-to-year scores — called value-added scores. The statistic used to evaluate value-added scores is the Cumulative Percent of Norm mean (CPN), which indicates the percent of the national (expected) gain attained. A CPN equal to 100 means that the student attained in that year, for his or her grade, in a particular subject, the gain equivalent to (i.e., 100 percent of) the national norm gain. To examine this concept further, let's consider a hypothetical example:

Suppose that in mathematics, the average national scale score on a standardized test is 550 in grade 4 and 600 in grade 5. Therefore, the national norm gain in math from grade 4 to grade 5 is 50 scale-score points. If Keisha, a fifth-grader, gains 75 points over her fourth-grade score, her CPN would be 150, because 75/50 = 1.50 x 100. Keisha thus has gained 150 percent of the national norm gain. Depending on her other scores on the standardized achievement test, she may have different CPNs for the remaining four subjects (language, reading, science, and social studies).

Suppose that we find Keisha's average CPN across the five subjects to be 125. The implication then is that over-

WHAT VALUE-ADDED "ADDS"

Value-added test results are very useful as a supplement to conventional school achievement data such as the percentile scores typically reported by the media. For example, the local newspaper may report that Highland Elementary School had an overall median percentile of 27 on this year's state mathematics test, while Brookline Elementary's median percentile was 55 on the same test. Thus, half of Highland's students scored below the 27th percentile for the state, while half of Brookline's students scored below the 55th percentile. Given that an average national median percentile would be 50, Highland doesn't appear to be doing well.

But what these results don't indicate are the types of students that Highland and Brookline are serving and how much progress is being made from year to year. Investigating this, we find that 98 percent of Highland's students and only 15 percent of Brookline's are eligible for free or reduced-price lunch. Further, Highland's median score last year was only the 18th percentile, while Brookline's was the 60th percentile. Although Highland appears to have made dramatic improvements this year, the newspaper data alone convey a very different story to the public.

What if the newspaper also reported Highland's average gain? Suppose, for example, the average gain per student at Highland were two times (i.e., 200 percent of) the national norm gain. The impression now is that even though Highland's students average well below national norms, they are bridging the gap rapidly. Maybe they've had a successful year with their CSR design after all!

all. Keisha has gained 125 percent of the national norm gain. She apparently has made good progress in her fifth-grade year. However, demonstrating strong gains doesn't mean necessarily that Keisha is demonstrating proficiency in a subject area. That is, based on the information given above, we don't know if Keisha's actual score relative to the scores of her classmates or relative to national norms is high or low. To understand how high Keisha actually is achieving, we need to know her actual score, her percentile score, and her gain score.

In the same way CPN scores are used to assess individual students' progress from year to year, CPNs can be calculated for entire schools based on their students' mean scores. For example, if School A had an average CPN gain of 100 percent in math, it would have achieved at the national (expected) level of achievement gain for that subject for that year. However, if its CPN in social studies were 80 percent, the school's students' average gain in that subject would have been only 80 percent of the expected gain.

Why value-added scores are useful:

- **They control for student mobility by excluding students who have not been in the district for successive years.** Median percentile scores typically report results for an entire school population. But if 50 percent of the students were transfers with limited exposure to the CSR design being evaluated, it would be unfair to include them in an assessment of design effects.
- **They control for student socioeconomic status by comparing students to themselves on year-to-year growth.** Suppose that a school's student population were becoming increasingly disadvantaged over time. The school's overall median percentile may decline steadily because of increased poverty, even though the CSR program may raise achievement for most individual students. Value-added scores will reveal such student gains.
- **They are sensitive to the progress of every student.** A school's median percentile may not change much from year to year, even though there is progress in raising student achievement. For example, if a school's median national percentile were 30, and many low-achieving students were moved from below the 15th percentile to close to the 25th percentile, the median percentile still would remain at 30, erroneously suggesting zero growth!
- **They demonstrate progress by an entire school or school district in raising student achievement up to standards.** Again, conventional measures like median percentiles may be insensitive to positive changes and thus lead to erroneous negative judgments about comprehensive school programs that are, in fact, narrowing the gap.

Key Findings From Memphis

So what do these results really mean? And how have the evaluation results helped Memphis make strategic decisions regarding the impact of CSR designs on schools? The results imply that the NAS designs were having a positive influence on student achievement after only two years of implementation. For supporters of CSR designs in Memphis, this came as very good news. Teachers, students, district personnel, and the CSR Design Teams now had solid, reliable data to support their requests for funding, flexibility, and innovation.

Value-added scores demonstrate progress by an entire school or school district in raising student achievement up to standards.

But the continuing challenge at all levels in Memphis is understanding how CSR designs are improving student learning, which elements of designs are most and least critical to improvements, and how such positive effects can be maintained and increased in subsequent years. A comprehensive evaluation program like the one used in Memphis can provide administrators and teachers with valuable information on how the CSR designs are bringing about improvements in the school building and community.

Detailed reports on the 34 Memphis schools that have implemented CSR designs can be found in such sources as Ross, Anderson et al. (1997); Ross, Troutman et al. (1997); Smith et al. (1998); and Stringfield et al. (1997). Here are their major conclusions:

- most schools have shown good year-to-year progress in implementing their CSR designs;
- elementary schools have made the fastest progress;
- student achievement has increased relative to achievement in schools that have not implemented CSR designs; and
- formative evaluation has provided data to enable schools to monitor their progress, be accountable to stakeholders, and engage in continuous improvement efforts.

BEST COPY AVAILABLE

Although not specifically addressed in this document, the Memphis evaluation also provided additional findings:

- teacher planning time has increased;
- most schools made appropriate selections of NAS designs based upon needs, student characteristics, teacher interests and skills, and other factors;
- teacher buy-in and principal leadership were highly critical to program success; and
- instructional strategies have changed to make student learning more active (using discussion, projects, cooperative learning, technology, and teacher coaching).

Using Evaluation for Continuous Improvement in Memphis

Evaluation needs to be an ongoing process, so the Memphis studies will continue. After a CSR program is established, there may be less need to repeat a comprehensive evaluation year after year. But as time passes, there will be new teachers, new administrators, different students, and likely changes in district or state policies. Continuous evaluation together with continuous improvement efforts can help ensure that a design is being implemented effectively and is producing the desired results. When shared with the school staff, evaluation results can be considered carefully in suggesting needed program refinements for the next year.

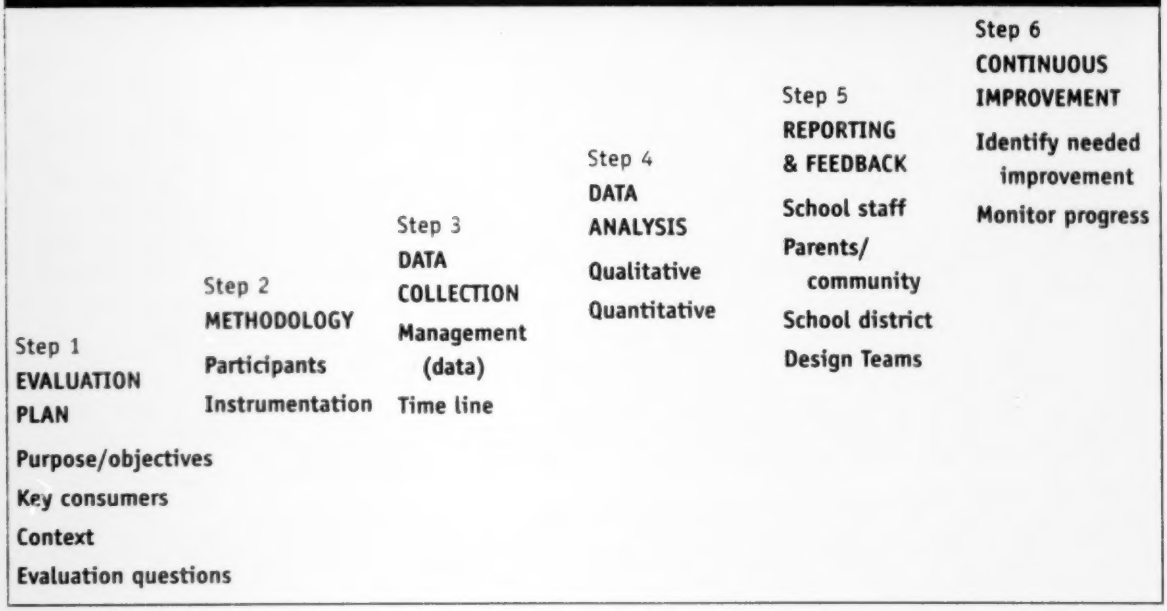
School districts have a critical role in the formative evaluation process. Without assistance from districts, many schools might fail to include formative evaluation as part of their reform efforts. The reasons are many: too much to do, too little time, a lean budget, limited expertise and experience, anxieties about evaluation, and so on. Schools shouldn't be forced to do evaluations, but they may need to be encouraged and assisted in obtaining needed information and resources. The school district, as evidenced in Memphis and other places, is best positioned and equipped to provide such assistance. Sharing the information in this guidebook should be an effective early step.

CONCLUSION

New American Schools sponsored the production of this guide to help districts, states, and schools successfully evaluate their CSR programs. Both formative and summative evaluations are an essential part of the reform process.

NAS encourages both school districts and schools to use these evaluation approaches for accountability purposes and for program improvement. The results should be communicated to serve these purposes not only internally and locally, but to NAS and the Design Teams as well. We strongly encourage the use of multiple assessments — both "process" (implementation) and "product" (outcomes) oriented — to help school staffs understand not only their degree of success but also the reasons for the outcomes attained. To assess student achievement, four levels of evaluation can be employed, the most rigorous of which involves the comparison of CSR program outcomes on both standardized and performance tests with outcomes of matched control schools. Reporting both value-added and normative achievement scores provides a more meaningful picture of achievement than relying on either score alone. For the broader purpose of improving American education, NAS further regards valid evaluation across numerous and diverse sites as essential to determining which CSR designs work most effectively in different contexts. We hope schools and districts find this publication valuable. We welcome your comments and feedback.

THE EVALUATION PROCESS — FIGURE 1



EXPEDITIONARY LEARNING OUTWARD BOUND FIGURE 2

Benchmarks, Indicators, and Evidence

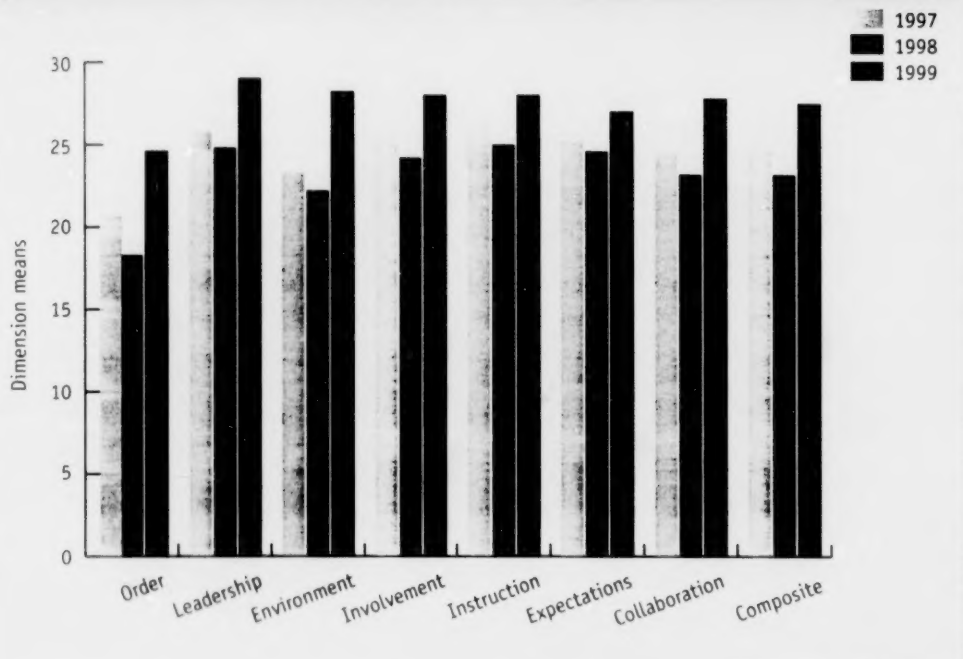
Benchmark	Phase I		Phase II		Phase III	
	Indicator	**Evidence**	**Indicator**	**Evidence**	**Indicator**	**Evidence**
II. Instruction *A. Expeditions* In a fully implemented Expeditionary Learning Outward Bound school, expeditions are the major part of the curriculum and include multiple disciplines, project work, exhibitions, portfolios, and a meaningful service component.	Teachers integrate some field work into expeditions.	• observation • teacher feedback	Expeditions include field work, multiple disciplines, exhibitions, and reflection.	• observation • documentation • teacher feedback	Expeditions include field work, multiple disciplines, exhibitions, reflection, and a meaningful service component.	• curriculum • documentation • teacher feedback
B. Peer Review and Revision In a fully implemented Expeditionary Learning Outward Bound school, peer review and revision are used on a daily basis as a means for perfecting the final product.	Students begin to do revisions and multiple drafts to get a final product.	• observation • teacher feedback	Students do multiple drafts, based on peer review, to develop a final product.	• observation • teacher feedback	Students use peer review daily as the basis for product drafts.	• observation • teacher feedback
	The teacher sets the standard criteria for the final product.	• observation • teacher feedback	Students begin to establish the criteria for the final product.	• observation • teacher feedback	Students agree on criteria for final product before work begins.	• observation • teacher feedback

Figure 2 shows a sample benchmarking instrument used by Expeditionary Learning Outward Bound, one of the NAS designs. Note that the benchmark for a particular design element is a generic statement about which events, activities, or structures will be included in a fully implemented design. More specific indicators and associated evidence then are stated for beginning (Phase I), intermediate (Phase II), and full (Phase III) implementation.

Developed by Memphis City Schools and the Center for Research in Education Policy, U of M

SPRING 1999 SCHOOL CLIMATE AUDIT FIGURE 3

Cloverdale Elementary School



TSCI Tennessee School Sample 10

COMPREHENSIVE SCHOOL REFORM (CSR) TEACHER QUESTIONNAIRE FIGURE 4
SUMMARY OF TEACHER PERCEPTIONS (EXTRACTED SAMPLE)
Cloverdale Elementary School

Spring 2000	Percent Agree	Percent Neutral	Percent Disagree
1. I have a thorough understanding of this school's comprehensive school reform (CSR) program.	40%	40%	20%
2. I have received adequate initial and ongoing professional development for CSR program implementation.	10%	25%	65%
3. The elements of our CSR program are integrated effectively to help us meet school improvement goals.	80%	15%	5%
4. Student achievement has been impacted positively by CSR.	70%	20%	10%
5. Our school has a plan for evaluating all components of our CSR program.	65%	35%	0%

Figure 4 illustrates the report format for teacher responses to items on the Comprehensive School Reform Teacher Questionnaire (CSRTQ) used in formative evaluation conducted jointly by the Center for Research in Educational Policy, the University of Memphis, and the AEL Regional Educational Laboratory, Charleston, W.V. The instrument contains 28 items that are reflective of CSRD criteria and is copyrighted by the Center for Research in Educational Policy.

SCHOOL OBSERVATION MEASURE (SOM) DATA SUMMARY FIGURE 5

For use in CSR formative evaluation conducted jointly by the Center for Research in Educational Policy, the University of Memphis, and the AEL Regional Educational laboratory, Charleston, W.V.

School name _____ Observer name _____

Date of observation _____ SOM _____ Observer role/affiliation _____

Number of classroom observations comprising this SOM _____

Directions: Use your class-specific notes to reflect upon the extent to which each of the following is present in the school:

	0 — Not Observed	1 — Rarely	2 — Occasionally	3 — Frequently	4 — Extensively
INSTRUCTIONAL ORIENTATION					
Direct instruction with the entire class (lecture)	❑	❑	❑	❑	❑
Team teaching	❑	❑	❑	❑	❑
Cooperative/collaborative learning	❑	❑	❑	❑	❑
Individual tutoring (teacher, peer, aide, adult volunteer)	❑	❑	❑	❑	❑
CLASSROOM ORGANIZATION					
Ability groups	❑	❑	❑	❑	❑
Multiage grouping	❑	❑	❑	❑	❑
Work centers (for individuals or groups)	❑	❑	❑	❑	❑
INSTRUCTIONAL STRATEGIES					
Instructional feedback (written or verbal) to enhance student learning	❑	❑	❑	❑	❑
Integration of subject areas (interdisciplinary/thematic units)	❑	❑	❑	❑	❑
Project-based learning	❑	❑	❑	❑	❑
Use of higher-level questioning strategies	❑	❑	❑	❑	❑
Teacher acting as coach/facilitator	❑	❑	❑	❑	❑
Parent/community involvement in learning activities	❑	❑	❑	❑	❑
STUDENT ACTIVITIES					
Independent seatwork (self-paced worksheets, individual assignments)	❑	❑	❑	❑	❑
Experiential, hands-on learning	❑	❑	❑	❑	❑
Systematic individual instruction (differential assignments geared to individual needs)	❑	❑	❑	❑	❑
Sustained writing/composition (self-selected or teacher-generated topics)	❑	❑	❑	❑	❑
Sustained reading	❑	❑	❑	❑	❑
Independent inquiry/research on the part of students	❑	❑	❑	❑	❑
Student discussion	❑	❑	❑	❑	❑
TECHNOLOGY USE					
Computer for instructional delivery (e.g., CAI, drill and practice)	❑	❑	❑	❑	❑
Technology as a learning tool or resource (e.g., Internet research, spreadsheet or database creation, multimedia, CD-ROM, laser disk)	❑	❑	❑	❑	❑
ASSESSMENT					
Performance assessment strategies	❑	❑	❑	❑	❑
Student self-assessment (portfolios, individual record books)	❑	❑	❑	❑	❑
SUMMARY ITEMS			*1 = Low, 2 = Moderate, 3 = High*		
Academically focused class time			❑	❑	❑
Level of student attention/interest/engagement			❑	❑	❑

RUBRIC FOR SOM SCORING

(0) Not Observed: Strategy was never observed.

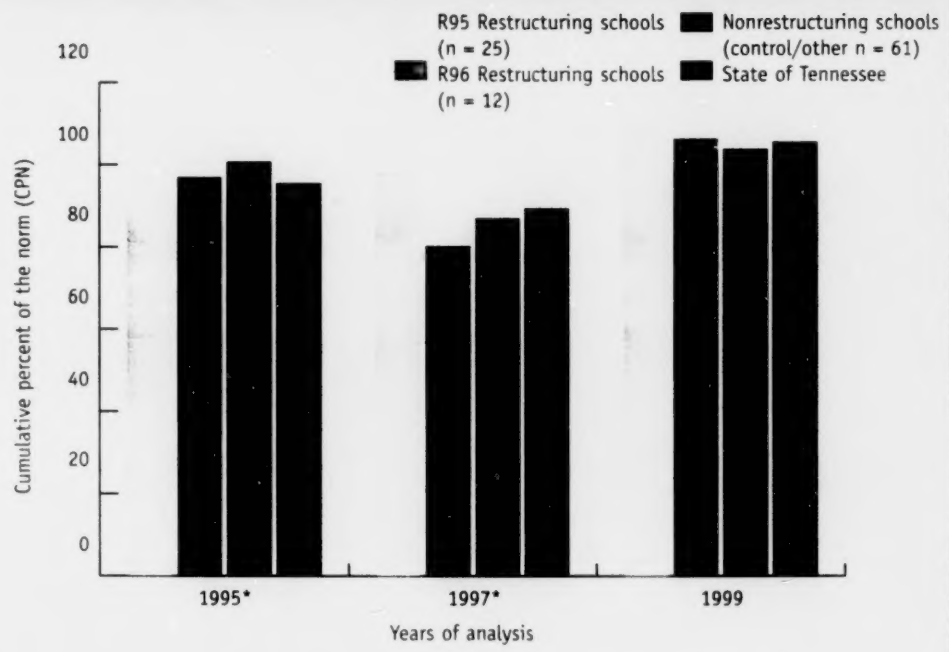
(1) Rarely: Observed in only one or two classes. Receives isolated use and/or little time in classes. Clearly not a prevalent/emphasized component of teaching and learning across classes.

(2) Occasionally: Observed in some classes. Receives minimal or modest time or emphasis in classes. Not a prevalent/emphasized component of teaching and learning across classes.

(3) Frequently: Observed in many but not all classes. Receives substantive time or emphasis in classes. A prevalent component of teaching and learning across classes.

(4) Extensively: Observed in most or all classes. Receives substantive time and/or emphasis in classes. A highly prevalent component of teaching and learning across classes.

MEMPHIS CITY SCHOOLS TVAAS RESULTS FOR ALL SUBJECTS AVERAGED FIGURE 6
(Grades 3–5)



*Note: 1995–97 are adjustments of actual TCAP scores based on TerraNova norms.

Figure 6 shows the TerraNova CPN means across several years for four categories of Memphis elementary schools: 25 schools that started restructuring in 1995 (R95), 12 schools that started restructuring in 1996 (R96), 61 nonrestructuring schools (NR), and all Tennessee schools. As shown in the figure, in spring 1995, before restructuring, the R95 schools (mean CPN = 91.5) were achieving lower scores than NR schools (mean CPN = 100.8) and all Tennessee schools as a group (mean CPN = 95.5).

In 1997, after almost two years of restructuring, R95 schools (mean CPN = 98.0) outperformed both NR schools (mean CPN = 87.1) and all Tennessee schools (mean CPN = 89.6). The R96 schools, during their first year, achieved relatively small gains (mean CPN = 80.3). In 1998, R95 (mean CPN = 108.3) and R96 schools (mean CPN = 106.5) showed slight performance increases over NR (mean CPN = 104.0) and all Tennessee schools (mean CPN = 105.7).

Looking at the 1997 results in Figure 6, we can see that R95 schools did well in comparison to all Tennessee schools and the NR schools. But why would gains for an entire state go down from 1995 to 1997? We feel certain that no statewide events (whether political, athletic, or weather-related) had a systemic negative influence on teaching or curriculum in all or most districts. However, each year the TerraNova test form changes, and as a result, some forms are more difficult relative to a particular state's or region's curriculum. The 1997 form was a much more difficult test than tests from previous years, but still, the R95 schools made gains relative to the "easier" 1995 test.

REFERENCES

Herman, R. and Stringfield, S. (April 1995). *The Promising Programs for Educating Disadvantaged Students: Evidence of Impact*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Ross, S. M.; Anderson, R.; Bassoppo-Moyo, T.; Bol, L.; Buggey, T.; Dietrich, A.; Grehan, A.; Hacker, D.; Henry, D.; Lowther, D.; Maxwell, S.; McNelis, M.; Nath, L.; Pace, J.; Phillipsen, L.; Rich, L.; Smith, L.; Troutman, A.; Wasson, R.; and Weddle, K. (1997). *Evaluation of New American Schools Designs: 1996–1997, Memphis*. Report prepared for Memphis City Schools. Memphis, TN: Center for Research in Educational Policy, the University of Memphis.

Ross, S. M. and Morrison, G. R. (1995). Evaluation as a Tool for Research and Development: Issues and Trends in Its Applications in Educational Technology (pp. 491–521). In R. D. Tennyson and A. Barron (Eds.), *Automating Instructional Design: Computer-Based Development and Delivery Tools*. New York, NY: Springer Verlag.

Ross, S. M.; Sanders, W. L.; Wright, S. P.; and Stringfield, S. (1998). *The Memphis Restructuring Initiative: Achievement Results for Years 1 and 2 on the Tennessee Value-Added Assessment System (TVAAS)*. Memphis, TN: Center for Research in Educational Policy, the University of Memphis.

Ross, S. M.; Smith, L. J.; Lohr, L.; and McNelis, M. (1994). Math and Reading Instruction in Tracked First-Grade Classes. *The Elementary School Journal, 95* (2), 105–119.

Ross, S. M.; Troutman, A.; Horgan, D.; Maxwell, S.; Laitinen, R.; and Lowther, D. (1997). The Success of Schools in Implementing Eight Restructuring Designs: A Synthesis of First-Year Evaluation Outcomes. *School Effectiveness and School Improvement, 8*, 95–124.

Ross, S. M.; Wang, W.; Sanders, W. L.; Wright, S. P.; and Stringfield, S. (June 1999). *Two- and Three-Year Achievement Results on the Tennessee Value-Added Assessment System for Restructuring Schools in Memphis*. Memphis, TN: Center for Research in Educational Policy, the University of Memphis.

Sanders, W. L. and Horn, S. P. (1995a). Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes. *Educational Policy Analysis Archives, 3* (6).

Sanders, W. L. and Horn, S. P. (1995b). The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment. In A. J. Shinkfield and D. Stufflebeam (Eds.), *Teacher Evaluation: Guide to Effective Practice* (pp. 337–350). Boston, MA: Kluwer.

Smith, L.; Ross, S.; McNelis, M.; Squires, M.; Wasson, R.; Maxwell, S.; Weddle, K.; Nath, L.; Grehan, A.; and Buggey, T. (1998). The Memphis Restructuring Initiative: Analysis of Activities and Outcomes That Impact Implementation Success. *Education and Urban Society, 30* (3), 296–325.

Stringfield, S. C.; Datnow, A.; Herman, R.; and Berkeley, C. (1997). Introduction to the Memphis Restructuring Initiative. *School Effectiveness and School Improvement, 8*, 3–35.

GLOSSARY OF TERMS

- **Benchmarking** is the process through which school staff reflect on, discuss, and document operationally the implementation goals of their comprehensive reform design. Benchmarks specify what will be achieved in early, intermediate, and full phases in the areas of curriculum, instruction, and organization.
- **Comprehensive School Reform** (CSR) covers all grade levels, all students, all subjects, and all functions of schooling (i.e., curriculum, instruction, standards, assessment, governance, and professional development).
- **Cumulative Percent of the Norm** (CPN) scores are used in the Tennessee Value-Added Assessment System to reflect the percentage of the national norm gain that a student attains in a given year for his or her grade in a particular subject. For example, if the national norm scale score gain from fourth to fifth grade in mathematics is 50 points, and Student A gains 50 points from fourth to fifth grade in math, then the student has gained exactly 100 percent of the national norm gain, and his or her CPN would be 100 percent.
- **Design-Based Assistance** (DBA) is a service offered by Design Teams to schools in the areas of training and professional development, materials and supplies, implementation checks and benchmarks, routine visits and support, and networking with other schools.
- **Formative Evaluation** involves assessing programs that are not yet completed or fully implemented. The primary purpose of formative evaluation is to obtain data to guide program improvement.
- **Performance Measures** assess students on what they know and are able to do, not just on what information they have learned. Such assessments typically require students to demonstrate learning on open-ended tasks by writing, presenting, performing, explaining, and exhibiting.
- **Summative Evaluation** involves assessing fully implemented programs to determine the degree to which they are satisfying their objectives. The primary purpose of summative evaluation is to obtain data to guide decisions about program effectiveness.
- **The Tennessee Value-Added Assessment System** (TVAAS) is a highly sophisticated system that applies complex mixed model statistical methodology to derive value-added scores (see Sanders & Horn, 1995a, 1995b) that control for both prior performance and socioeconomic status. In other words, a high (or low) CPN score is equally attainable by poor and wealthy students and by high and low achievers. School districts probably cannot duplicate a TVAAS-level system on their own (i.e., without considerable statistical help or involvement by Sanders and colleagues), but they certainly can look at students' and schools' changes in scores from year to year in addition to the usual median percentile data.
- **Value-Added Assessment** is a measure of the degree to which a program or intervention adds "value" or gain for recipients. In assessing student achievement, value-added scores reflect how much students improve on a standardized test in particular subjects from one year to the next.

BEST COPY AVAILABLE

Berends, Mark (1999). *Assessing the Progress of New American Schools: A Status Report*. Santa Monica, CA: RAND.

Berends, Mark and Sheila Kirby (1999). *Implementation and Performance in New American Schools: An Executive Summary*. Santa Monica, CA: RAND.

Berends, Mark, with Sheila Kirby, Scott Naftel, and Christopher McKelvey (1999). *Implementation and Performance in New American Schools: Three Years After Scale-Up*. Santa Monica, CA: RAND.

Bodilly, Susan (1996). *Lessons From New American Schools Development Corporation's Demonstration Phase*. Santa Monica, CA: RAND.

Bodilly, Susan (1998). *Lessons From New American Schools' Scale-Up Phase: Prospects for Bringing Designs to Multiple Schools*. Santa Monica, CA: RAND.

Bodilly, Susan, with Susanna Purnell, Kimberly Ramsey, and Christina Smith (1995). *Designing New American Schools: Baseline Observations on Nine Design Teams*. Santa Monica, CA: RAND.

Cicchinelli, Louis F. and Zoe Barley (1999). "Evaluating for Success." *Comprehensive School Reform: An Evaluation Guide for Districts and Schools*. Aurora, CO: McREL.

Glennan, Thomas (1998). *New American Schools After Six Years*. Santa Monica, CA: RAND.

Keltner, Brent (1998). *Funding Comprehensive School Reform*. Santa Monica, CA: RAND.

Mitchell, Karen (1996). *Reforming and Conforming: NASDC Principals Discuss School Accountability Systems*. Santa Monica, CA: RAND.

Maria Voles Ferguson served as project manager for New American Schools.

Editorial and production assistance was provided by KSA Group, Inc.

New American Schools Designs

As of February 2000, New American Schools (NAS) is at work in more than 2,100 schools around the country. NAS district partners commit to transforming a minimum of 30 percent of their schools within five years. Most partners are on track to meet and exceed this goal by year three. The eight NAS designs are listed below.

America's Choice School Design

Formerly known as the National Alliance for Restructuring Education, America's Choice is built on a framework of high academic standards and matched assessments. It incorporates a standards-based curriculum focused on the basics, conceptual mastery, and applications. The design quickly identifies students who fall behind and brings them back to standard, and includes a planning and management system for making the most efficient use of available resources to raise student performance.
For more information: 202-783-3668;
e-mail: schooldesign@ncee.org; www.ncee.org.

ATLAS Communities

The ATLAS design centers on pathways — groups of schools made up of high schools and the elementary and middle schools that feed into them. Teams of teachers from each pathway work together to design curriculum and assessments based on locally defined standards. The teachers in each pathway collaborate with parents and administrators to set and maintain sound management and academic policies, ultimately resulting in improved student performance.
For more information: 617-969-7100;
e-mail: Atlas@edc.org; www.edc.org/FSC/ATLAS.

Co-NECT Schools

Assisting schools in creating and managing their own high-tech equipment and network, Co-NECT uses technology to enhance every aspect of teaching, learning, professional development, and school management. Co-NECT schools are organized around small clusters of students who are taught by a cross-disciplinary team. Most students stay in the same cluster for at least two years. Teaching and learning revolve around interdisciplinary projects that promote critical skills and academic understanding, as well as integrate technology.
For more information: 617-873-5612;
e-mail: info@co-nect.com; www.co-nect.com.

Expeditionary Learning Outward Bound

Built on 10 design principles, Expeditionary Learning Outward Bound (ELOB) operates on the belief that learning is an expedition into the unknown. ELOB draws on the power of purposeful, intellectual investigations — called learning expeditions — to improve student achievement and build character. Learning expeditions are long-term, academically rigorous, interdisciplinary studies that require students to work inside and outside the classroom. In ELOB schools, students and teachers stay together for more than a year, teachers work collaboratively, and tracking is eliminated.
For more information: 617-576-1260;
e-mail: info@elob.org; www.elob.org.

Modern Red Schoolhouse Institute

This design strives to help all students achieve high standards through the construction of a standards-driven curriculum, use of traditional and performance-based assessments, establishment of effective organizational patterns and professional development programs, and implementation of effective community-involvement strategies. Students master a rigorous curriculum, develop character, and promote the principles of democratic government. These elements of the traditional red schoolhouse are combined with a high level of flexibility in organizing instruction and deploying resources, use of innovative teaching methodologies, student groupings for continuous progress, and advanced technology as a learning and instructional management tool.
For more information: 888-275-6774;
e-mail: skilgore@mrsh.org; www.mrsh.org.

Roots and Wings

This elementary school design builds on the widely used Success For All reading program and incorporates science, history, and mathematics to achieve a comprehensive academic program. The premise of the design is that schools must do whatever it takes to make sure all students succeed. To this end, Roots and Wings schools provide at-risk students with tutors, family support, and a variety of other services. While the "roots" of the design refer to mastery of basics, the "wings" represent advanced accomplishments that students achieve through interdisciplinary projects and a challenging curriculum provided by the design.
For more information: 800-548-4998;
e-mail: rslavin@inet.ed.gov; www.successforall.net.

Turning Points

This middle school design focuses on creating a professional, collaborative culture to improve teaching and learning for adolescents. Turning Points schools commit to a multiyear, systemic change process and engage in practices that guide faculty collaborations and lead to significant and sustained improvement in student learning. Member schools are provided with a variety of supports including on-site coaching; professional development and networking; the Turning Points Self Study Survey, which helps schools identify challenges and set priorities; publications and technology; and accountability processes and assessment of student learning.
For more information: 617-421-0134;
e-mail: leah_rugen@ccebos.org.

Urban Learning Centers

The Urban Learning Centers (ULC) design is a comprehensive K–12 model for urban schools. The curriculum and instruction are designed to ensure that all students are taught in a K–12 community, enabling new strategies to overcome barriers by addressing the health and well-being of students and their families. Governance and management also are restructured to engage community members in decisionmaking and to ensure that the design can improve and evolve. ULC also incorporates the extensive use of advanced technology as an essential element for implementation of the design.
For more information: 213-622-5237;
e-mail: gpruitt@laedu.lalc.k12.ca.us; www.lalc.k12.ca.us.

New American Schools

Papers in the *Getting Better by Design* series include ...

- **Design-Based Assistance as a Cornerstone of a School Improvement Strategy**
- **How to Create and Manage a Decentralized Education System**
- **How to Rethink School Budgets to Support School Transformation**
- **How to Rebuild a Local Professional Development Infrastructure**
- **How to Make the Link Between Standards, Assessments, and Real Student Achievement**
- **How to Create Incentives for Design-Based Schools**
- **How to Build Support for Comprehensive School Reform**
- **How to Evaluate Comprehensive School Reform Models**

Accompanying this series are New American Schools Action Tools

To help you implement the ideas and suggestions recommended in the *Getting Better by Design* series, New American Schools is creating hands-on Action Tools that complement and expand the use of the research papers. As they become available, each tool will be posted on the NAS Web site, www.naschools.org.

For more information about the *Getting Better by Design* series ...

For more information about the *Getting Better by Design* series and the corresponding Action Tools, or to obtain copies of the *Getting Better by Design* "How-To" papers, write to New American Schools, 1560 Wilson Boulevard, Suite 901, Arlington, VA 22209 or call 703-908-9500. NAS also can be reached by e-mail at info@nasdc.org or via the World Wide Web at www.naschools.org.

Education Commission of the States

This publication was made possible, in part, from funding received from the Education Commission of the States (ECS) through a generous grant from the Annenberg Foundation. ECS's role as a partner in the New American Schools effort is to support national dissemination of the NAS designs and to work with state policymakers to create the policy changes necessary to help the designs flourish.